

The Quality of Ministerial Accountability in the Canadian House of Commons

Tanya Whyte*

*Department of Political Science, University of Toronto
Receptiviti Inc.*

Dated: June 3, 2019

Abstract

A core tenet of responsible government is that people authorized to act are accountable for how they use that authority. Each civil servant has a boss, who has a boss, and so on. Authority runs top-down from ministers to deputies to assistant deputies and downward to further subordinates. Accountability runs bottom-up from subordinates to bosses and finally to ministers. Ministers have final authority because they are ultimately accountable, individually and collectively, to elected representatives in the House of Commons. Authority begins where accountability ends. In theory.

What is the quality of accountability in the House of Commons? How can it be measured? How does it vary over time and across governments? How is it affected by ministerial experience, government popularity, or time to election? This paper draws on key conceptualizations of accountability, new tools in Natural Language Processing, and the newly digitized corpus of Parliamentary Debates (www.lipad.ca) to examine systematically the properties and quality of ministerial answers to opposition questions in the Canadian House of Commons. The results have implications for the theory of responsible government, discursive institutionalism, and computational political science.

*Email: tanya.whyte@mail.utoronto.ca. Paper prepared for the conference of the Canadian Political Science Association, Vancouver, BC, June, 2019. This paper is a draft. Do not cite without permission.

1 Introduction

How accountable is Canadian democracy? The answer to this question is bound up with the concept of responsible government, the basic principle of our Westminster-style parliamentary system. Governments are responsible and accountable to parliament, and therefore to the people. A core tenet of responsible government is that anyone authorized to act is accountable for how they use that authority. Authority runs top-down from ministers to deputies to assistant deputies and downward to further subordinates. Accountability runs bottom-up from subordinates to bosses and finally to ministers. Ministers have final authority because they are ultimately accountable, individually and collectively, to elected representatives in the House of Commons. Those representatives—in practice, opposition members—are tasked with holding ministers, including the Prime Minister, to account. Conventionally, members of the House of Commons possess the institutional power to withhold their confidence in the government’s authority, provoking its dissolution. This check on the government’s power is rarely used, and, in practical terms, is only possible during a minority parliament.

Instead, the day-to-day legitimation of political authority in the Canadian House of Commons takes place through dialogue and debate. Ideally, this collective performance forces the government to provide information and rationales for its decisions, gives oppositions a public platform to demonstrate their fitness to govern, and raises public consciousness of important policy issues. Question Period is the most visible and dramatic example of debate as an accountability mechanism. It offers a daily, 45 minute opportunity for Opposition members to ask questions of any government minister, including the Prime Minister. It has also been widely criticized for its inefficacy throughout its institutional history (Chong, Jennings, Laframboise, Davies, & Lukiwiski, 2010; Docherty, 2005; Loat & MacMillan, 2014; Stewart, 1977).

This paper is an empirical assessment of the quality of ministerial accountability in Question Period. I leverage the dialogic nature of responsible government to conceptualize and measure accountability by using the text of parliamentary speeches as quantitative data. Whyte (2019) proposes such a measurement approach based upon a theoretical model that draws on the comparative literature on parliamentary debate. Parties in the House of Commons face a trade-off between using their allocated time to take an ideological stand or to engage on matters of accountability. Institutional rules and norms shape a balance between these two goals moderated by the current state of a political incentive structure. This structure is affected by variables like seat count, polling popularity, internal discipline, and time to the next election. Two mathematically related but conceptually distinct measures of textual similarity can be employed to assess the relative importance of accountability or ideology in a party’s speeches. The extent to which two speeches are semantically similar (that is, they share the same meanings) is employed as a measure of their ideological similarity. The extent to which texts are lexically similar (or share the same words) is a measure of the quality of accountability in a debate forum.

This paper focuses on using the latter measure to test predictions about ministerial accountability generated within the above theoretical framework. More specifically, I extend the methodology and analysis developed in Whyte (2019) to investigate whether incorporating information about ministerial portfolios into the model of parliamentary accountability provides additional analytical leverage. Using a subset of the Lipad dataset of Canadian parliamentary debates (Beelen et al., 2017), I generate a study dataset of individual Question Period interactions between government ministers and oppositions, enriched with additional metadata about ministers, debate topics, and parliaments. I look for evidence of the following propositions: are minority parliaments more accountable than majority parliaments? Is the quality of accountability affected by the government’s popularity, or the time until the next election? Finally, does accountability vary at the individual minister level according to cabinet portfolio?

2 Theory

This paper draws on previous quantitative work in the comparative literature on the empirical study of legislative debate. Such work is primarily grounded in a rational choice perspective, which emerged as an adaptation of models of voting behaviour, particularly spatial models of roll call voting in legislatures, to speeches. Proksch and Slapin (2014) distill this perspective into a formal behavioural model. Their basic assumption is that legislators make speeches in order to claim a position on an issue and communicate it to voters and to their parties. Parties are interested in presenting a united electoral front to voters, and thus are incentivized to assert control over their members’ speeches. Individual representatives face a choice between toeing the party line or defecting in order to cultivate a personal vote in their district. Proksch and Slapin argue that electoral rules are what shape this decision; the more parties benefit electorally from cultivating a united brand, the more strongly they will assert control over individual speechmaking (Proksch & Slapin, 2014, 27–28).

In a Westminster-style parliament, however, legislative power is institutionally entangled with executive control, and parties are interested in maintaining discipline for other reasons apart from electoral success. Bäck and Debus (2016) propose an alternative model that envisions an individual legislator’s decision to speak as shaped by three types of potential gain: collective political goals, individual political goals, and individual selective incentives. Collective political goals include electoral success (or office-seeking), but may also include, for example, policy-seeking on the part of a party that is a member of a governing coalition. Parties’ histories, ideological commitments, and leaders’ personal characteristics may also shift the weight they place on one collective goal over another. Exemplifying the debt of this literature to rational choice models of voting behaviour, Bäck and Debus propose a “calculus of speechmaking“ based on the probability that a given speech will be decisive in attaining some set of collective and individual goals (Bäck & Debus, 2016, 29). The very low probability of making such a decisive vote is why the calculus of voting model elaborated

by Riker and Ordeshook (1968) makes use of a controversial “duty” term representing an individual’s intrinsic motivation to vote in order to balance expected costs and benefits of voting. In the adaptation of this model to speeches, the costs are even higher and benefits more nebulous. However, Bäck and Debus’ model provides a path forward for incorporating institutional variables other than electoral system into a model of parliamentary speechmaking.

For example, the enforcement of party discipline can serve a collective aim above and beyond what is optimal for a given electoral institution. This is a particularly important insight for understanding the Canadian case. Proksch and Slapin’s model predicts loose party discipline should reign in a Westminster-style parliament combined with a first-past-the-post electoral system, like Britain’s, as the incentive for an MP to cultivate a personal vote in his or her district is strong. However, this is clearly not the case in Canada, where defection in roll-call votes and in speeches is consistently very rare. In the Canadian House of Commons, the dilemma of whether or not an MP should defect against their party, or whether a party leader should decide not to enforce discipline, is rarely an interesting empirical problem to study. For example, party leadership simply does not delegate control over the decision to speak to any individual MP, a practice that has been institutionalized in the conventional list of speakers the Speaker calls upon during Question Period (Bosc & Gagnon, 2017).

Instead, to understand the Canadian case, we can look to other institutional variables that shape the incentive structure of parties to speak *about* particular topics. Following Whyte (2019), this paper proposes an alternative model combining Proksch and Slapin’s focus on the meaningfulness of speech content with Bäck and Debus’ understanding of multiple and potentially competing collective goals. The relative strength of a government in terms of seats controlled, the potential for strategic cooperation with third parties, and the strength of party discipline and caucus unity are all examples of internal institutional variables that affect the strategic calculus of parties in Parliament (Gervais, 2012, 10).

The approach taken here is to relax the restrictions imposed by a formal representation of individual decision-making, and think instead at the collective level of the party as parliamentary actor. In general, I propose that ideology and accountability are the two main collective goals of debate, serving purposes that can be framed as, broadly speaking, policy-seeking and office-seeking. An alliance of opposition parties facing a tenuous minority government, but lacking the popularity to attain office themselves, may face some likelihood of success at forcing policy concessions out of the government. Or, a strong majority government nearing the end of its term may choose to focus on presenting its accountability to voters and address the skeletons in its closet in the House rather than deal with them unexpectedly on the campaign trail.

Assume an opposition party receives some opportunity to ask questions based upon their relative level of representation in the House of Commons, and will use that opportunity to most effectively further the political goals they value the most. The goal they choose to emphasize depends on factors that influence the likelihood of success of attaining either of these goals. Under typical majority conditions, for example, an opposition party is likely to emphasise

office-seeking, lacking the institutional power to force policy concessions out of the government. They will instead attempt to make the government accountable for its decisions in a public fashion, uncover mismanagement, and present themselves as a viable alternative to hold office in the next election, whenever that may be.

The fundamental source of variation in this model is the threat of institutional sanction faced by the current government. This is a related but broader conceptualization of the role of the electoral system played in Proksch and Slapin's model. In other words, a government changes its accountability behaviour based upon shifts in its likelihood of losing power. Such possibilities include a vote of non-confidence in a minority parliament situation, or low polling popularity combined with a term nearing its end. The greater the threat of loss of office to a government, the more likely it will emphasize office-seeking in its speechmaking, to communicate to voters and to the House its fitness and legitimacy to govern. As discussed previously, however, parliamentary accountability is a dialogic process. The quality of accountability in the House of Commons as an institution also depends upon the collective incentives opposition parties, especially the Official Opposition, face. An opposition may choose to hold government to account or communicate its own ideological goals. As a practical example, an opposition that asks nothing but loaded or unanswerable questions contributes to impoverishing accountability in the same way as a Cabinet that refuses to answer (albeit to a lesser extent, of course, given the government's institutional power).

Drawing on this theory, I generate the following propositions for study in this paper. First, the threat of sanction implies that the quality of accountability should increase the more tenuous a government's control of the House. Therefore, I expect minority parliaments to be more accountable than majorities. Second, an alternative measure of threat to a government's viability is its polling popularity. In literal terms, this is the public's answer to the pollster's question of "If an election were to be held tomorrow, which party would you vote for?" The lower a government's popularity, the more likely it will face a confidence challenge (in the minority case) or will be unable to call an election when conditions are favourable (in the majority case). Thus, I expect that the lower the current polling popularity of the government, the higher the accountability measured in the House of Commons. Third and related, the influence of polls on a government's safety is mediated by not only minority status but by duration of government. The longer a minority government has endured, the more likely it will continue to stay in power; likewise, a majority government has the least to fear for its accountability early in its mandate. Thus, I expect that parliaments in their first session will be less accountable than those in later sessions. Fourth, the balance of collective effort towards accountability or ideology is likely to vary significantly over issue area and therefore ministerial portfolio: some policy portfolios are more ideologically-salient than others, and tactics may change accordingly.

3 Methodology

Assuming such a model of parliamentary speech, how can the relative balance between goals be measured, and therefore, the quality of accountability in the House of Commons? In this paper, the translation from theory to measurement builds on Bovens’ work on the conceptualization and measurement of accountability (Bovens, 2005, 2010). Bovens defines accountability as a social relation between an actor (in this case, the government) and an accountability forum (the House of Commons—or, in practice, its opposition members) with three stages. First, the actor is obliged to explain and justify his or her exercise of power to the forum by providing relevant information and rationales. Second, the forum has the opportunity to question the actor and interrogate the information provided. Third, the forum is empowered to pass judgement on the actor and impose consequences if deemed necessary (Bovens, 2005, 184–186). According to Bovens’ model, what would effective accountability look like in the House of Commons? We would expect to observe a regular cycle of government MPs providing information and opposition MPs asking questions. Potentially, if the opposition were unsatisfied with the government’s data and explanations, it could withdraw confidence in the government.

The threat of institutional sanction, as discussed above, affects the willingness of the government to candidly provide information. It also affects the opposition’s motivation to spend time on interrogating complex issues, versus, for example, scoring partisan points in Question Period. If the opposition asks substantive questions and the government responds with relevant information—in other words, in a case of high accountability—both sides will make some shared reference to the same data, rationales, regulations, or policy proposals under discussion. In this case, I expect the government and opposition to overlap significantly in the words they use in reference to these shared discussion points. To use a term from computational linguistics, their speeches on a given shared topic will have high lexical similarity.

Lexical similarity, and, more concretely, machine learning models that exploit raw word frequencies as features, are commonly employed in the literature on legislative debate as a metric to measure ideological polarization. Peterson and Spirling (2018), for example, use the relative predictive performance of a text classification model to gauge the polarization of Conservative and Labour MPs in the United Kingdom. The foundational assumption of such approaches is that ideology is the main source of systematic variation in the words legislators select. In Peterson and Spirling’s model, words are envisioned as belonging to “left” or “right” buckets. Assuming a two-party system with one party on the right and one on the left, it is easiest to predict which party delivered a randomly-sampled speech if the left party consistently uses only left words and the right party only right words. The extent to which a left party uses right words and vice versa introduces error into the classification and potentially can be used to measure ideological overlap.

However, any speech will necessarily use words with no partisan meaning. Both parties may use these words at different rates, potentially confusing our

ability to distinguish between shared word use reflective of ideological similarity and shared word use that, in this context, constitutes random noise. Peterson and Spirling categorize such non-partisan into a third “noise“ category. They make the caveat that their ideological classification measure performs best when such noise is low and non-systematic, and perform diagnostic tests in their article to investigate this possibility. However, Hirst et al. (2010) find that government or opposition status is a persistent and major confound for ideological classification particularly in the Canadian House of Commons.

The amount of noise introduced by the attack and defence of parliamentary opposition, especially in the Canadian case, can be viewed as a troublesome obstacle to using words as data to measure ideological polarization. However, this paper proposes an alternative viewpoint: treating this noise as a dependent variable worthy of study rather than a source of error illuminates the dynamics of parliamentary accountability. In this paper, I focus on using lexical similarity to understand questions and answers in Question Period. For the more specific task of using words to measure ideology, Whyte 2019 proposes an alternative semantic similarity measure employing word embeddings that is not discussed in this paper.

Returning to Bovens’ model, the theoretical stages of parliamentary accountability should be most visible in the real world in the institution of Question Period. In the Canadian House of Commons since the mid-1970s, a formal 45-minute period has existed on the daily timetable during which opposition party members can freely question government ministers. I leverage this institutional feature to construct a dataset of question-answer exchanges between opposition members and government ministers. Beginning with transcripts from the Lipad database of House of Commons debates filtered for Question Periods from 1975-2010, I categorize speeches per day per debate subtopic. Then, within each of these daily-level subtopics, I identify pairs of speeches consisting of an opposition speech followed by a speech by a government minister. I calculate the lexical similarity between these two speeches, then average all such similarity scores for the day’s Question Period to constitute one observation. This method of dataset generation is a methodological variation on Whyte (2019), which concatenates government and opposition speeches per Question Period for calculation of a similarity score between two larger aggregates. It adapts the data generation method used in Whyte (2019) for qualitative validation of the accountability measure to a quantitative study, with the primary intention of assessing whether ministerial portfolios significantly affect quality of accountability. For more discussion of qualitative validation of a lexical similarity measure of accountability, including a case study of sample texts, please see Chapter 5 of (2019).

Following dataset generation, the calculation approach employed here is identical to that used in Whyte (2019). First, I use a simple hashing vectorizer (a more memory-efficient approach to count vectorization) to construct a term-document matrix and standardize these vectors with L2 normalization. Prior to vectorization, I perform some standard text preprocessing methods including stemming, lowercasing, and removal of numbers and punctuation. However, I

do not remove stopwords nor transform the term-document matrix using an algorithm such as tf-idf to reduce the influence of frequently-used words on the results. This decision is rooted in findings from the language psychology literature that coordination in use of function words across individuals and groups is reflective of lower social distance (Niederhoffer & Pennebaker, 2002). Finally, I calculate cosine similarities between the two speeches in each paired exchange, then take the mean of these observations per Question Period. I use a hierarchical mixed effects model to study the predictions made above including random effects at the parliament level to account for historical political variation such as changes in the party system. For a more extensive methodological and technical discussion, please see Chapter 4 of Whyte (2019).

4 Results and Discussion

To provide a historical overview of our results, Figure 1 shows the distribution of Question Period lexical similarity scores per Parliament. Visually, it is apparent that at the parliamentary level, accountability scores do not show any significant trend and are subject to substantial variance. Looking more closely at the Question Period-level data, results from an analysis of the three theoretical propositions outlined earlier in this paper are summarized in Table 1.

First, are minority parliaments more accountable than majorities? I do not find significant support for this proposition using this dataset. This finding runs contrary to the results in Whyte (2019), which found significant variation at the aggregated Question Period level across majority and minority parliaments. Second, does the current popularity of a government (measured using quarterly polling data) affect the level of accountability in the House of Commons? In this case, I confirm a similar finding to that in Whyte (2019): as a government becomes more popular, accountability goes down. In terms of effect size, the poll effect is the dominant fixed effect in the model ($R^2 \approx 0.1$ out of a model marginal $R^2 = 0.124$). Third, are governments in their first session less accountable than those currently on to their second or third? Since minority parliaments are more likely to survive only one session, the interaction between majority status and session is of particular interest. Majority governments are more accountable if in their second or later session, an effect that almost attains significance at $p = 0.076$. However, this effect is very small ($R^2 \approx 0.005$) and the overall result is weak enough to be discarded.

In contrast to the fixed effects, most of the variance captured by the model is attributable to the random effects component ($ICC = 0.53$, Conditional $R^2 = 0.591$). This phenomenon was also the case in Whyte (?), highlighting the importance of political variation across parliaments in interpreting any textual measure based on parliamentary speech data. The fact that many of the key institutional variables, such as balance of seats, important to determining a government's perceived threat to power vary significantly only from parliament to parliament casts doubt upon the approach using a dataset of individual-level minister exchanges as observations.

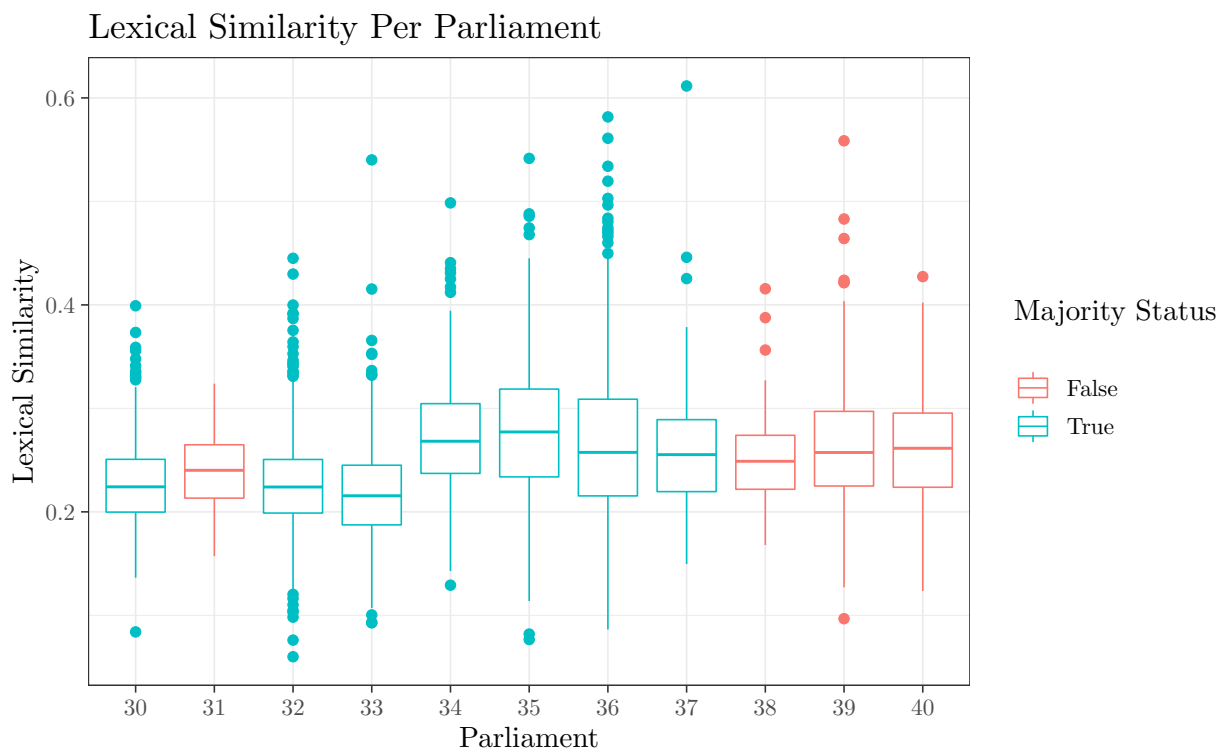


Figure 1: Lexical similarity per Question Period within each Parliament since 1975

Predictors	Lexical Similarity		
	Estimates	CI	p
(Intercept)	0.271	0.242–0.299	< 0.001
majorityTrue	-0.019	-0.048–0.010	0.223
firstsessionFalse	-0.007	-0.026–0.012	0.458
partyLiberal	0.004	-0.022–0.030	0.79
pollpercent	-0.001	-0.001– -0.000	0.044*
majorityTrue×firstsessionFalse	0.019	-0.002–0.039	0.076
Random Effects			
σ^2			0.00
$\tau_{00parliament}$			0.00
ICC _{parliament}			0.53
Observations			114
Marginal R^2 / Conditional R^2	0.124 / 0.591		

Table 1: Mixed-effects model of lexical similarity in Question Period per quarter

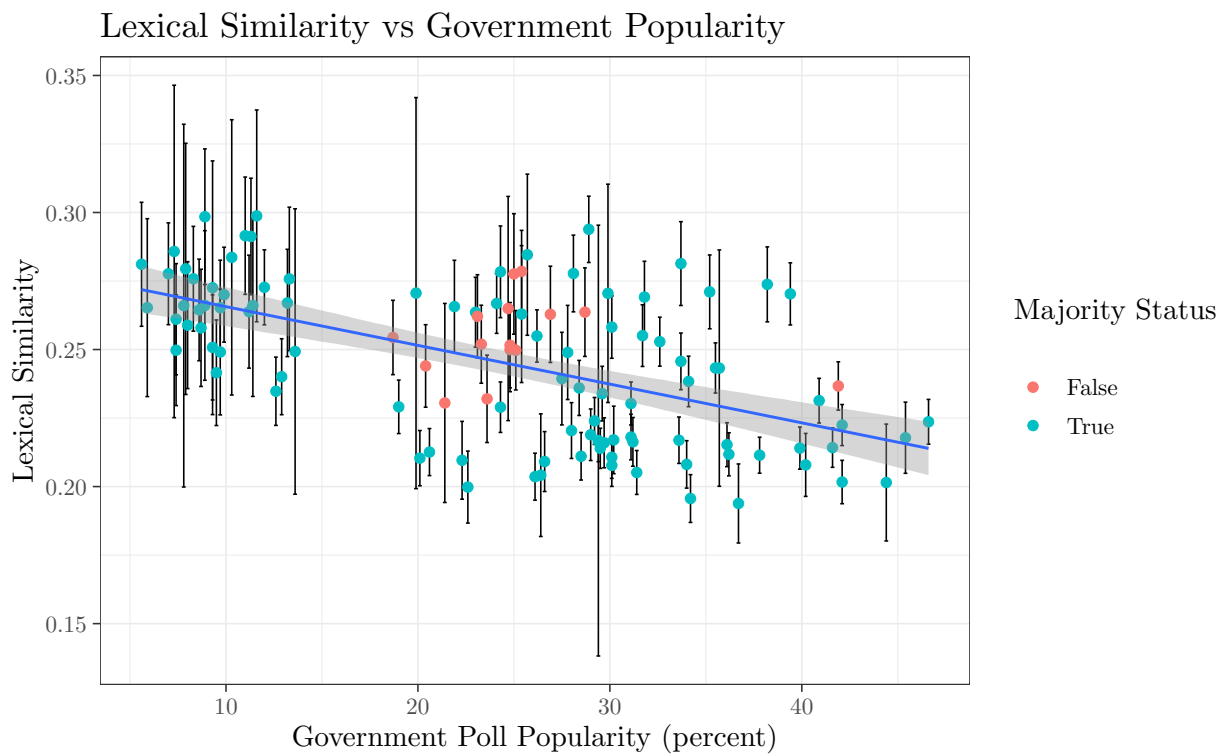


Figure 2: Polling popularity of the government (per quarter) versus mean lexical similarity per Question Period per quarter

In Figure 3, I select for display purposes only those ministerial portfolios that participated in at least one-fifth of the Question Periods in the dataset—in practical terms, those who spoke at least once a week on average, and participated in $n \geq 870$ debates. As can be seen visually in Figure 3, there is no apparent trend in accountability scores across types of ministerial portfolio. A complete model accounting for all existing levels of minister portfolio (123 types in the dataset overall) finds a small number of portfolios attain significance (such as Minister of Intergovernmental Affairs). However, bearing in mind the problem of multiple comparisons, the low effect sizes observed for these significant levels, and the fact that none of the “major” ministerial portfolios attained a significant effect, I conclude there is no significant evidence that ministerial portfolio has a significant effect on the quality of accountability in paired minister exchanges.

To summarize, I find that the methodology of dataset construction tested in this paper as an alternative to the approach in Whyte (2019) was less analytically useful than anticipated. One explanation for this outcome is a disconnect between the methodological approach of measurement at the individual MP level and the concept of collective goal-seeking that matters most in my theoretical model. These results could reflect the institutional distinction between collective accountability and individual ministerial responsibility, suggesting that an alternative approach to theorizing the latter might be necessary.

Alternatively, my experiment of narrowing the study dataset to match more closely the idealized empirical model of accountability could have introduced additional error into the accountability measure through omission of relevant speeches. Within the approach applied here, longer passages of discussion on the same topic are sometimes not pooled together but treated as separate observations of paired exchanges. The source of this inconsistency is the variable specificity with which Hansard records the subtopic of a given passage of debate in Question Period. As an illustrative example, a successive set of questions on May 8, 1975 were posed to the Minister of Finance on the subject of inflation. These question-answer pairs had subtopics THE CANADIAN ECONOMY, INFLATION - KNOWLEDGE OF MINISTER OF WORKING PAPERS PRESENTED TO CANADIAN LABOUR CONGRESS SEEKING CONSENSUS ON RESTRAINT, INFLATION - REQUEST FOR TABLING OF WORKING PAPERS SUBMITTED TO CANADIAN LABOUR CONGRESS, INFLATION - INQUIRY WHETHER GUIDELINES ON RESTRAINT HAVE BEEN PREPARED FOR ALL SECTORS OF THE ECONOMY, and INFLATION - POSSIBILITY OF MONITORING BOARDS WITH POWER TO ENFORCE GUIDELINES. Because of the different subtopic headings, my methodology classifies each of these exchanges as independent. However, in a real dialogue participants may refer back to questions or concepts discussed earlier to elaborate their answers or to seek further clarification. To get a more realistic understanding of the government’s accountability on inflation on this day, we would want to pool together all the opposition’s questions, and government answers, on inflation-related subjects. This fact informs the methodology employed in Whyte (2019), in which I aggregate all questions and all answers per Question Period into two texts for comparison. Nevertheless, it would be worth trying a

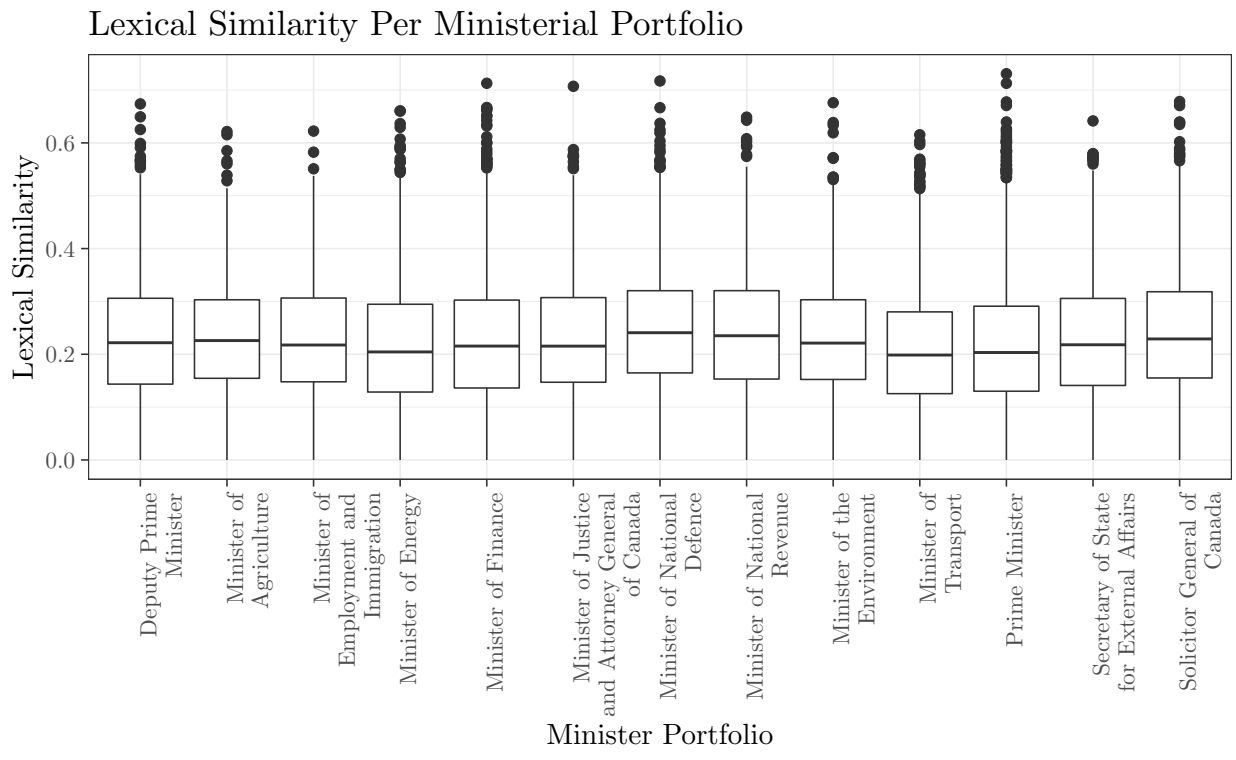


Figure 3: Lexical similarity per Question Period for the most frequently-occurring ministerial portfolios in the dataset.

middle of the road approach based using topic modelling or clustering to identify which questions and answers should be pooled each Question Period based on topical similarity. A comparative analysis of parliamentary debate datasets from other countries with a parliamentary Question Period testing these multiple methods of dataset generation would also illuminate this question.

Finally, despite the negative finding in my study of differences in accountability across ministerial portfolios, there are additional possibilities to study regarding individual ministerial responsibility. Following Bäck and Debus' model, there may be a distinction in personal style or incentives across MPs as individuals rather than as ministers of a given department. Studying MPs with long careers in Cabinet may reveal the extent to which these individual terms affect accountability. However, the limited pool of appropriate individuals to study, the overwhelming influence of party discipline in the Canadian case, and other confounding factors like the historical dominance of the Liberal party imply that a comparative analysis would be necessary to investigate this possibility.

References

- Bäck, H., & Debus, M. (2016). *Political Parties, Parliaments and Legislative Speechmaking*. New York: Springer.
- Beelen, K., Thijm, T. A., Cochrane, C., Halvemaan, K., Hirst, G., Kimmins, M., ... Whyte, T. (2017). Digitization of the Canadian parliamentary debates. *Canadian Journal of Political Science*, 50(3), 849–864.
- Bosc, M., & Gagnon, A. (2017). *House of Commons Procedure and Practice* (2nd ed.). Ottawa: Editions Yvon Blais.
- Bovens, M. (2005). The concept of public accountability. In E. Ferlie, L. E. Lynn, & C. Pollitt (Eds.), *The Oxford Handbook of Public Management* (pp. 182–208). New York: Oxford University Press.
- Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics*, 33(5), 946–967.
- Chong, M., Jennings, M., Laframboise, M., Davies, L., & Lukiwiski, T. (2010). What to do about Question Period: A roundtable. *Canadian Parliamentary Review*, 33(3), 2–8.
- Docherty, D. C. (2005). *Legislatures*. Vancouver: UBC Press.
- Gervais, M. (2012). *Challenges of Minority Governments in Canada*. Ottawa: Invenire Books.
- Hirst, G., Riabinin, Y., & Graham, J. (2010). Party status as a confound in the automatic classification of political speech by ideology. In *Proceedings, 10th International Conference on Statistical Analysis of Textual Data/10es Journées internationales d'Analyse statistique des Données Textuelles (JADT 2010), Rome* (pp. 731–742).
- Loat, A., & MacMillan, M. (2014). *Tragedy in the commons: Former Members of Parliament speak out about Canada's failing democracy*. Toronto: Random House Canada.

- Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*(4), 337–360.
- Peterson, A., & Spirling, A. (2018). Classification accuracy as a substantive quantity of interest: Measuring polarization in Westminster systems. *Political Analysis, 26*(1), 120–128.
- Proksch, S.-O., & Slapin, J. B. (2014). *The Politics of Parliamentary Debate*. Cambridge: Cambridge University Press.
- Riker, W. H., & Ordeshook, P. C. (1968). A theory of the calculus of voting. *American Political Science Review, 62*(1), 25–42.
- Stewart, J. B. (1977). *The Canadian House of Commons: Procedure and Reform*. Montreal: McGill-Queen's University Press.
- Whyte, T. (2019). *Quantitative measurement of parliamentary accountability using text as data: the Canadian House of Commons, 1945-2015* (Unpublished doctoral dissertation). University of Toronto.